Proc. of International Symposium on Managing Water Supply for Growing Demand
, Bangkok, 16-20, October, 2006, IHP Technical Documents in Hydrology, no. 6, pp. 97-108, 2006

# Stochastic Modeling of Rainfall Maxima Using Neyman-Scott Rainfall Model

**Carlo Mondoñedo, Yasuto Tachikawa, Kaoru Takara**
*Disaster Prevention Research Institute, Kyoto University, Uji, Kyoto 611-0011, Japan*
mondonedo@flood.dpri.kyoto-u.ac.jp

## ABSTRACT

When discharge records are inadequate in length, synthetic rainfall can be generated from stochastic techniques to supply crucial flood control decision variables (following the rules of rainfall-runoff modeling) in design and operation of a flood control structure. The Neyman-Scott Poisson Rectangular Pulse Rainfall Model was applied here to generate synthetic rainfall maxima. Applying this model required historical moments of the rainfall in the study areas. Several test sets, consisting of different numbers of historical moments were prepared to estimate NSM parameters for the regional rainfall of Kamishiiba (Kyushu), Naha (Okinawa), and Sapporo (Hokkaido). Based on these NSM parameters, random numbers of different types were generated from the inverse CDF method to generate the synthetic rainfall of each area. The resulting maxima were adequate for the majority of cases except when a mix of rainfall sources was prominent in several months of the year. In general, an ideal set of parameters was determined but the methodical testing of combinations of moments employed in the proposed test sets was recommended for future applications in regions elsewhere.

## Introduction

Flood control decision variables connected to design and operation depend on an appropriate availability of rainfall records. Ideally, such decision variables should be determined based on a quantile of flood discharge. Unfortunately, discharge records may be short (less than 30 years) and/or nonstationary due to land use change and/or river intrusive construction. In this case, synthetic rainfall time series generation coupled with rainfall-runoff modeling proves to be a valuable tool. For the general purpose of the former, stochastic techniques can be employed to generate synthetic rainfall and thus, synthetic quantiles. The generation of such synthetic rainfall records is thus an effective decision-making aid to water resources engineers.

Several stochastic methods under the point process (Cox and Isham, 1980) approach are available for synthetic rainfall generation. Such synthetic data can then be used in design storm evaluation for small retaining and impoundment structures as well as sewer systems (Cowpertwait, 1996). For such purposes, several families of stochastic models are available such as the Independent Poisson Marks Model (IPMM) (Eagleson, 1972), Poisson Rectangular Pulse Models (PRPM) (Rodriguez-Iturbe *et al.*, 1987), and Clustered Poisson Rectangular Pulse Models (CPRPM) (Rodriguez-Iturbe *et al.*, 1987 and Burlando and Rosso, 1993). The Neyman-Scott Poisson Rectangular Pulse Rainfall Model (NSM here, for brevity) (Rodriguez-Iturbe, 1987), a model from the latter type, was the key technique used in this study.

All previously mentioned models were based on the theory of point processes. Essentially, in point processes, probabilities can be mapped to random occurrences of point events, rainfall in this case. For instance, both the IPMM and PRPM initiate rainfall arrivals as Poisson occurrences (see Burlando and Rosso, 1993, for an example). Random variables such as duration and intensity of rainfall are considered exponential in distribution. The two models differ by the overlapping possible to the rectangular pulses of PRPM otherwise absent in IPMM. Applications of these models have appeared in the literature (Rodriguez-Iturbe *et al.*, 1987 and Burlando and Rosso, 1993) with the short-coming that the models developed could not be consistent with more than one aggregation period.

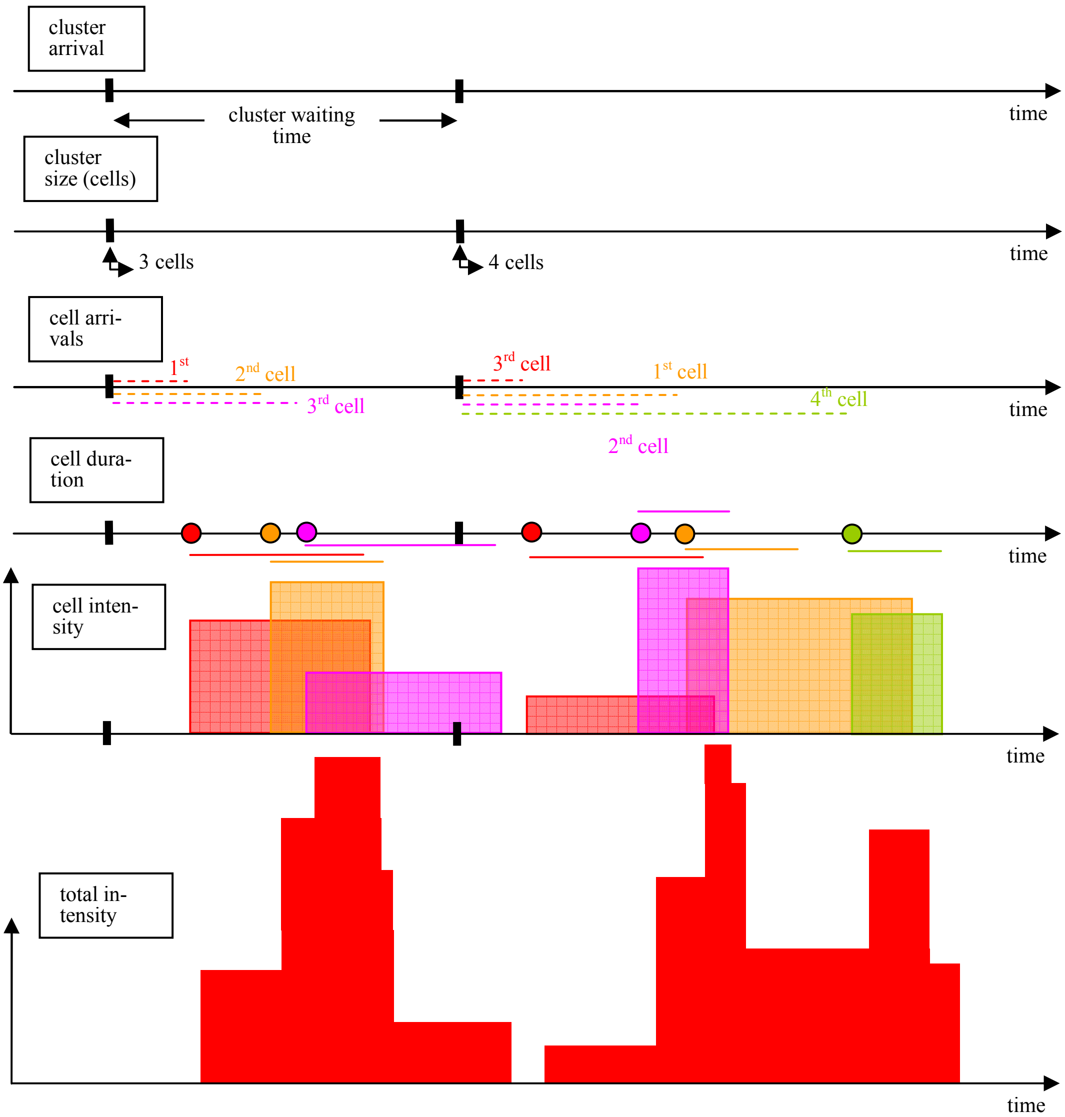


Fig 1. Schematic diagram of the Neyman-Scott Model.

The models under the category of CPRPRM are an application of point processes whereby the synthetic data generated can be consistent with more than one aggregation level (i.e.: synthetic hourly and daily rainfall is consistent with historical counterparts). In essence, rainfall arrives in clusters of random arrival time and number, each cluster consisting of rain cells of random birth, intensity, and duration. In addition, the Neyman-Scott model (NSM) is a CPRPRM in which rain cells arrive subsequent to the arrival of a cluster's arrival.

Unlike previous studies, the emphasis here is to investigate the ability of the NSM to preserve the historical quantile rainfall depths of 1-hour and 24-hour duration, which may be basic information necessary for risk analysis in flood control decision-making. In so doing, longer NSM rainfall records, although synthetic, can be available as reliable bases of quantile events.

The historical data used in this study were obtained from three locations to incorporate the effect on the quantiles of rainfall generated by fronts and/or typhoons. Sixteen yearly records (1988 to 2003) of hourly rainfall were taken from Kamishiiba (Kyushu Island), while 26 years (1976-2002) each of hourly rainfall were taken from Naha (Okinawa) and Sapporo (Hokkaido).

**Neyman-Scott Poisson Rectangular Pulse Rainfall Model**

Figure 1 show the random processes involved in the concept of the Neyman-Scott model. This basic version consists of essentially five probability distributions. In this NSM, clusters of cells are linked integrally to a storm origin with mean occurrence rate λ, regarded as a Poisson process, where waiting times between clusters are exponential in λ. The arrivals of these clusters are shown in the first time line of Fig. 1. Each storm can have a random number of cells described by a geometric distribution (with all clusters containing at least one cell), as shown in the second time line. Relative to the cluster origin, the random arrival of each cell is based on an exponential distribution, as shown in the third time line. Each cell has a corresponding independent identically distributed (iid) random intensity and duration, also based on the exponential distribution, shown in the fourth and fifth time line, respectively. The total rainfall intensity is then the superposition of the effects of these random cell intensities, as shown in the sixth time line. A succinct representation of the previously mentioned distributions can be written as:

$$p(N = n) = \frac{v^n e^{-v}}{n!} \quad (1)$$

$$f(t_s) = 1/\lambda \exp(-\lambda t_s) \quad (2)$$

$$p[C = c] = \frac{(1 - 1/\mu_c)^{c-1}}{\mu_c} \quad (3)$$

$$f(t_d) = \beta \exp(-\beta t_d) \quad (4)$$

$$f(i_c) = 1/\mu_x \exp(-1/\mu_x i_c) \quad (5)$$

$$f(t_c) = \delta \exp(-\delta t_c) \quad (6)$$

where:
$v$ = mean number of occurrences
= λT; T is the time period in consideration
$\lambda$ = mean arrival rate of a storm
$p[N=n]$ = probability that the number of clusters $N$ is equal to $n$
$t_s$ = storm arrival time
$f(t_s)$ = probability that the arrival of a storm origin is $t_s$
$p[C=c]$ = probability that the number of cells of a storm $C$ is equal to $c$
$\mu_c$ = mean number of cells in a storm
$f(t_d)$ = probability that the arrival of a cell from the storm origin is $t_d$
$1/\beta$ = mean displacement of a cell from the storm origin
$f(i_c)$ = probability that the intensity of a cell is equal to $i_c$
$\mu_x$ = mean intensity of a cell
$f(t_c)$ = probability that the duration of a cell's life is equal to $t_c$
$1/\delta$ = mean cell life span

Inherent in the model is the assumption of stationarity in the mean and variance. In applying the model therefore, it would be beneficial to have as long a historical rainfall record as possible. Based on the method of moments, the historical rainfall record can be expressed in terms of the model parameters as (Rodriguez-Iturbe, 1987):

$$E\langle Y_i^{(h)} \rangle = \lambda(\mu_c)h/(\delta\mu_x) \quad (7)$$

$$\mathrm{var}\langle Y_i^{(h)} \rangle = \frac{\lambda(\mu_c^2 - 1)[\beta^3 A_1(h) - \delta^3 B_1(h)]}{\beta\mu_x^2\delta^3(\beta^2 - \delta^2)} + \frac{4\lambda\mu_c A_1(h)}{\mu_x^2\delta^3} \quad (8)$$

$$\mathrm{cov}(Y_i^{(h)}, Y_{i+k}^{(h)}) = \frac{4\lambda\mu_c A_2(h,k)}{\mu_x^2\delta^3} + \frac{\lambda(\mu_c^2 - 1)[\beta^3 A_2(h,k) - \delta^3 B_2(h,k)]}{\beta\mu_x^2\delta^3(\beta^2 - \delta^2)} \quad (9)$$

in which:

$$A_1(h) = \delta h - 1 + e^{-\delta h}$$

$$B_1(h) = \beta h - 1 + e^{-\delta h}$$

$$A_2(h,k) = 0.5(1 - e^{-\delta h})^2 e^{-\delta h(k-1)}$$

$$B_2(h,k) = 0.5(1 - e^{-\beta h})^2 e^{-\beta h(k-1)}$$

where:
$i$ = time interval counter
$h$ = integer specifying time step interval of data (1 for 1 hour, 24 for 1 day, etc.)
$Y_i^h$ = rainfall depth in the i-th time of interval h

$E\left\langle Y_i^{(h)} \right\rangle$ = mean rainfall depth record at h-hours

$\operatorname{var}\left\langle Y_i^{(h)} \right\rangle$ = variance of rainfall record at h-hours

$\operatorname{cov}\left(Y_i^h, Y_{i+k}^h\right)$ = covariance or rainfall record at h-hours at lag k

**Parameter Estimation**

Although the model parameters can be estimated by maximum likelihood, this study followed previously adopted techniques from the method of moments. Indeed, a major limitation of the former method is the lack of the features of aggregated rainfall over a certain interval of time from the sample data sets, which was the case in this study. Hence, the method of moments was adopted here instead.

Five parameters are required in the NSM: λ, δ, $\mu_c$, $\mu_x$, and β. As shown previously, these parameters are directly connected to sample moments of the historical rainfall records in the form of equations (7)-(9). Several nontrivial combinations of these equations are used to form systems to be solved numerically. Normally, these systems are solved for the required parameters by minimization of an objective function. In this study, the following objective function was used.

$$F = \sum_{j=1}^{M} \left( \frac{f_j(Y_i)}{W_j} - 1 \right)^2 \qquad (10)$$

where:

$f_j(Y_i)$ = j-th NS moment equation of rainfall depth $Y_i$ (from equations (7) –(9)).

$W_j$ = historical moment value from rainfall record.

$M$ = number of equations to be adopted in the estimation.

Table 1. Proposed test sets for NSM parameter estimation problem.

| Test Set | Moments | Hours of Aggregation to be Used | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 6 | 12 | 24 | 48 |
| I | Mean | o | | | | |
| | Variance | o | | | o | |
| | Covariance* | o | | | o | |
| II | Mean | o | | | | |
| | Variance | o | o | | o | |
| | Covariance* | o | o | | o | |
| III | Mean | o | | | | |
| | Variance | o | | o | o | |
| | Covariance* | o | | o | o | |
| IV | Mean | o | | | | |
| | Variance | o | | | o | o |
| | Covariance* | o | | | o | o |
| V | Mean | o | | | | |
| | Variance | o | o | o | o | |
| | Covariance* | o | o | o | o | |
| VI | Mean | o | | | | |
| | Variance | o | | o | o | o |
| | Covariance* | o | | o | o | o |

*Covariance at lag 1

This choice of (10) was made here to ensure that large numerical values do not dominate the fitting procedure (Favre *et al.*, 2004). To apply (10), one must adopt a combination of equation (7) – (9), depending on the target use of the resulting NSM. Such combinations include those used in the studies of Rodriguez-Iturbe *et al.* (1987), Burlando and Rosso (1993), Cowpertwait *et al.* (1996), Calenda and Napolitano (1999), and Favre *et al.* (2004), to cite a few. These combinations range from the most basic (5 equations in the objective function to solve for the 5 parameters are used) to the more thorough (more than 5 equations in the objective function to solve for the 5 parameters are used). For instance, the determination the five parameters of the NS model can include the following equations, designated here as Test Set I:

1.hourly mean of rainfall depth ((7) cast in h = 1 hr)
2.variance of hourly rainfall depth ((8) cast in h = 1 hr)
3.variance of daily rainfall depth ((8) cast in h = 24 hrs)
4.lag-1 covariance of hourly rainfall depth ((9) cast in h = 1 hr, and lag k = 1)
5.lag-1 covariance of daily rainfall depth ((9) cast in h = 24 hr, and lag k = 1)

For this study, the hourly and daily maxima of the synthetic data should be sufficiently close to those of the historical sample. Since the above-mentioned set does not directly employ these maxima in the estimation of the parameters, it would be appropriate to assume that Test Set I may not cover this requirement for all possible rainfall conditions (i.e.: the temporal storm structure of rainfall may vary by location and season). To cover possible dependencies of the target maxima on the short-term and long-term moments of the historical data (if such dependencies exist), six combination test sets were adopted, as shown in Table 1.

The minimization technique adopted here included a means to employ the constraints shown in Table 2. Napolitano and Calenda (1999), in their study on unbiased parameter estimates for the NSM, adopted these ranges. The application of these ranges in this current study include the use of the Nelder-Mead Simplex (Press, *et al.*, 1992) and Levenberg-Marquardt (Press, *et al.*, 1992) minimization techniques in tandem for initial and refined estimation, respectively.

Table 2. Range of NSM parameters used for optimization.

| Parameter | Min | Max |
|---|---|---|
| $\lambda$ (1/h) | 0.001 | 0.050 |
| $\mu_c$ | 2.0 | 100.0 |
| b (1/h) | 0.01 | 0.50 |
| $\mu_c$ (mm/h) | 0.30 | 15.0 |
| $\eta$ (1/h) | 0.10 | 5.0 |

## Random Number Generation

To use the NSM parameters, a uniform deviate generator (random numbers within (0,1)) was developed from the method described in Press *et al.* (1992) which made use of the Park-Miller "minimal standard" generator based on the simple multiplicative congruential algorithm:

$$I_{j+1} = \xi I_j \left(\mathrm{mod}\ \ m\right) \tag{11}$$

where:

$\xi$ = multiplier = $7^5$ = 16,807
$m$ = modulus = $2^{31} - 1$ = 2,147,483,647
mod = modulus operator
$I_j$ = previous random integer between 0 and $m$-1
$I_{j+1}$ = succeeding random integer between 0 and $m$-1.

The inverse CDF method (ICDFM) was used for generating continuous random variables while a look-up table implementation of this method was used for generating discrete random variables (see Gentle, 1998).

## Application of the Neyman-Scott Rectangular Pulse Rainfall Model

Parameters were estimated on a monthly basis in what was considered the season of heavy rainfall of June to October. This was done to maintain the stationarity assumption of NSM. Three locations were selected to consider the effects of rainfall in a typhoon-frequented area (Naha, Okinawa), a front-frequented area (Sapporo, Hokkaido), and both (Kamishiiba, Kyushu). Parameters obtained from each set for each month of each area were used to generate synthetic rainfall with the exception of Sapporo, where Test Sets IV and VI did not seem to yield reasonable parameter estimates (see discussion on Sapporo below).

One hundred synthetic records were generated for each parameter set of each month. Synthetic moments were calculated from each set and compared with the historical counterparts. For each test of each month, a search for monthly hourly and daily maxima was conducted, enabling a Kolmogorov-Smirnoff (KS) test between historical maximum and synthetic maximum. The KS tests used here was based on the set of equations (12) – (13). The KS probability here is a variant of the original KS test that uses the Kuiper statistic, $V$, defined in this application as the sum of the maximum distance of a cumulative frequency distribution function of the synthetic maxima above and below the cumulative frequency distribution of the historical maxima, yielding a more sensitive test at the extreme ends of the CDF.

$$Q_{KS}(\gamma) = 2\sum_{j=1}^{\infty}\left(4j^2\gamma^2 - 1\right)e^{-2j^2\gamma^2} \tag{12}$$

$$Q_{KS}(0)=1 \quad ; \quad Q_{KS}(\infty)=0 \qquad (13)$$

The value of the KS probability $P_{KS}$ is then given in equation (14). High values $P_{KS}$ disproves the null hypothesis; the two samples compared may originate from the same population.

$$P_{KS}(V > observed) = Q_{KS}\left(\left[\sqrt{N_e} + 0.155 + 0.24/\sqrt{N_e}\right]D\right) \qquad (14)$$

$$N_e = \frac{N_h N_s}{N_h + N_s} \qquad (15)$$

where:
$D$ = maximum deviation between cumulative frequency distributions of historical and synthetic rainfall maxima.
$N_h$ = length of historical record.
$N_s$ = length of synthetic record.

This version of the KS Test was taken from Press, et al (1992).

***Parameter Estimates***

In general, the value of the objective function (10) was quite higher when it was adopted in estimation of NSM parameters in Test IV, V, and VI. It was observed that the correlation coefficient of the historical data dropped from about 0.6 at the original hourly level to around ±0.1 at the 48-hourly aggregated data. This weaker correlation structure of the historical data at the 48-hour level seemed to affect the results of Test IV and Test VI more than the Test V case. Appendix 1 and 2 contain partial historical moments and NSM parameter estimates obtained, respectively.

The moments of the synthetic record were compared to the historical records by calculating the residual in equation (16). Ideally, for each synthetic record, one must obtain a value as close to zero as possible for the hourly and daily data. Generally, the residual of the synthetic daily records were larger than the residual of the synthetic hourly records, as shown in Table 3.

$$R_l = \sum_{j=1}^{3}\left(1 - \frac{SM_{l,j}}{HM_{l,j}}\right)^2 \qquad (16)$$

where:
$R_l$ = residual of the l-th hourly aggregated data
$SM_{l,1}$ = Mean of synthetic rainfall record at the l-th aggregation level
$SM_{l,2}$ = Variance of synthetic rainfall record at the l-th aggregation level
$SM_{l,3}$ = correlation at lag-1 of synthetic rainfall record at the l-th aggregation level
$HM_{l,1}$ = Mean of historical rainfall record at the l-th aggregation level
$HM_{l,2}$ = Variance of historical rainfall record at the l-th aggregation level
$HM_{l,3}$ = correlation at lag-1 of historical rainfall record at the l-th aggregation level

Table 3. Residuals of synthetic moments to historical moments

| KAMISHIIBA | JUNE | | JULY | | AUGUST | | SEPTEMBER | | OCTOBER | |
|---|---|---|---|---|---|---|---|---|---|---|
| **TEST** | **HOURLY** | **DAILY** | **HOURLY** | **DAILY** | **HOURLY** | **DAILY** | **HOURLY** | **DAILY** | **HOURLY** | **DAILY** |
| I | 0.000 | 0.006 | 0.000 | 0.001 | 0.001 | 0.002 | 0.001 | 0.007 | 0.017 | 0.020 |
| II | 0.001 | 0.004 | 0.000 | 0.001 | 0.001 | 0.001 | 0.001 | 0.007 | 0.023 | 0.001 |
| III | 0.005 | 0.007 | 0.002 | 0.005 | 0.004 | 0.004 | 0.004 | 0.005 | 0.013 | 0.026 |
| IV | 0.013 | 0.427 | 0.001 | 0.042 | 0.003 | 0.005 | 0.000 | 0.016 | 0.001 | 0.001 |
| V | 0.000 | 0.025 | 0.008 | 0.020 | 0.003 | 0.021 | 0.001 | 0.005 | 0.039 | 0.168 |
| VI | 0.016 | 0.080 | 0.003 | 0.037 | 0.001 | 0.051 | 0.001 | 0.002 | 0.026 | 0.057 |
| **NAHA** | **JUNE** | | **JULY** | | **AUGUST** | | **SEPTEMBER** | | **OCTOBER** | |
| **TEST** | **HOURLY** | **DAILY** | **HOURLY** | **DAILY** | **HOURLY** | **DAILY** | **HOURLY** | **DAILY** | **HOURLY** | **DAILY** |
| I | 0.012 | 0.031 | 0.006 | 0.006 | 0.001 | 0.002 | 0.012 | 0.013 | 0.004 | 0.011 |
| II | 0.012 | 0.010 | 0.006 | 0.050 | 0.005 | 0.012 | 0.003 | 0.004 | 0.004 | 0.008 |
| III | 0.003 | 0.004 | 0.005 | 0.064 | 0.012 | 0.015 | 0.001 | 0.004 | 0.007 | 0.029 |
| IV | 0.012 | 0.032 | 0.006 | 0.017 | 0.011 | 0.046 | 0.001 | 0.001 | 0.012 | 0.034 |
| V | 0.005 | 0.018 | 0.003 | 0.153 | 0.004 | 0.027 | 0.002 | 0.006 | 0.002 | 0.005 |
| VI | 0.006 | 0.257 | 0.003 | 0.052 | 0.034 | 0.159 | 0.001 | 0.025 | 0.002 | 0.011 |
| **SAPPORO** | **JUNE** | | **JULY** | | **AUGUST** | | **SEPTEMBER** | | **OCTOBER** | |
| **TEST** | **HOURLY** | **DAILY** | **HOURLY** | **DAILY** | **HOURLY** | **DAILY** | **HOURLY** | **DAILY** | **HOURLY** | **DAILY** |
| I | 0.243 | 0.234 | 0.001 | 0.002 | 0.001 | 0.095 | 0.005 | 0.083 | 0.002 | 0.002 |
| II | 0.230 | 0.173 | 0.011 | 0.160 | 0.001 | 0.012 | 0.001 | 0.031 | 0.002 | 0.018 |
| III | 0.241 | 0.354 | 0.006 | 0.002 | 0.001 | 0.004 | 0.000 | 0.006 | 0.001 | 0.017 |
| IV | -not applied- | | -not applied- | | -not applied- | | -not applied- | | -not applied- | |
| V | 0.005 | 0.026 | 0.001 | 0.016 | 0.003 | 0.054 | 0.005 | 0.052 | 0.001 | 0.043 |
| VI | -not applied- | | -not applied- | | -not applied- | | -not applied- | | -not applied- | |

*Kamishiiba, Kyushu (1988-2003)*

The Kyushu Region exhibits mixes of rainfall sources, although typhoons are more erratic in occurrence in the summer. Several rainfall moments are shown in Appendix 1, taken from the Kamishiiba Observatory in Kyushu. As expected, the residuals between synthetic and historical moments for this area are quite small, shown in Table 3. This indicates the ability of NSM to preserve the historical moments in the generated synthetic rainfall. This applies to all the months of the rainy season in the area. This was the original objective in the conceptualization of the model and was considered (in this study) secondary to the effectiveness of the model to yield synthetic maxima similar to the historical maxima.

It was predetermined that a 95% value of the KS probability was a practical value for disproving the null hypothesis (although other tests can be employed). Thus, at 95% KS probability, the maximums of synthetic data appear to come from the same population as those of the historic data. The results of hourly and daily synthetic maxima of the study areas were grouped into three categories (see Figure 2):

(1) Both hourly and daily maxima display a KS probability $P_{KS} \geq 95\%$.

(2) Both hourly and daily maxima display a KS probability $P_{KS} < 95\%$.

(3) Only the hourly or daily maxima display a KS probability $P_{KS} \geq 95\%$.

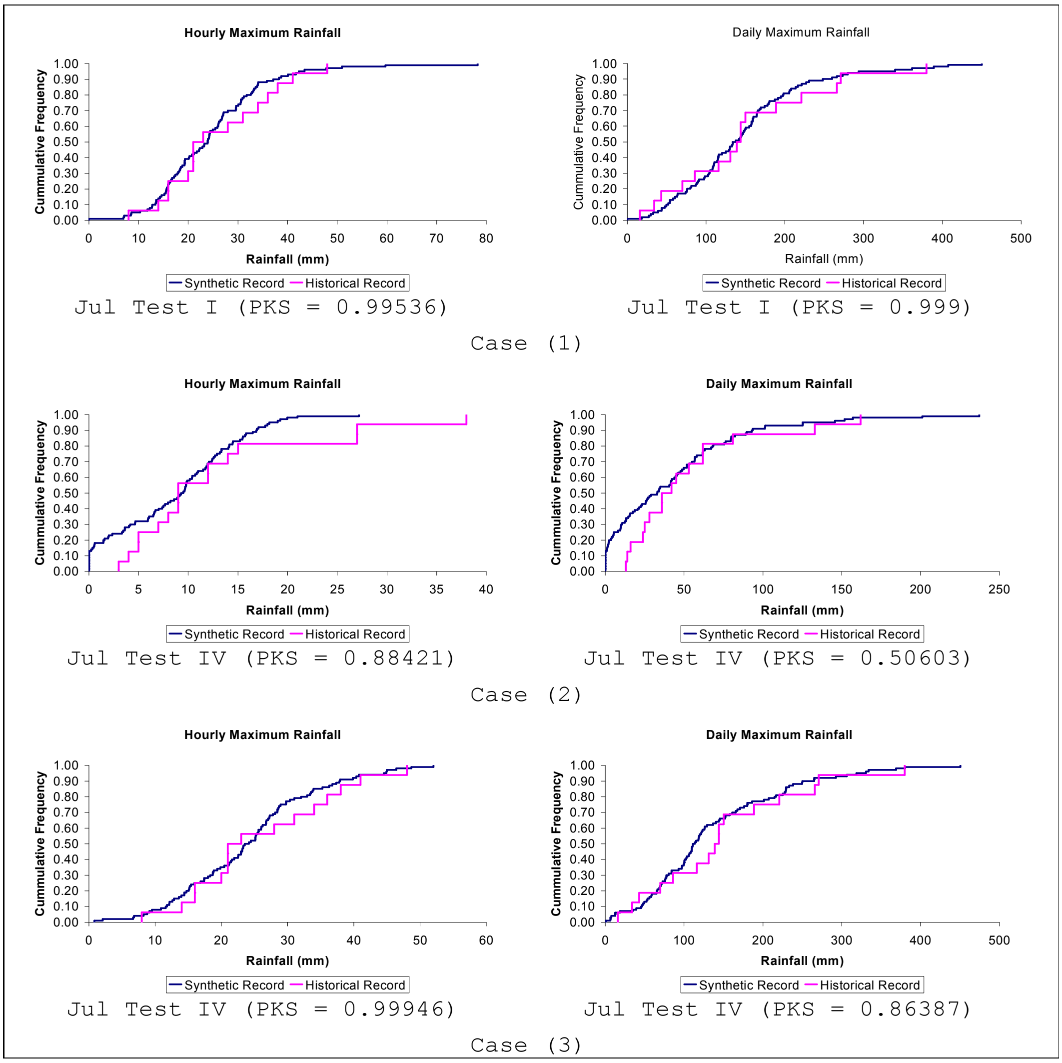


Fig. 2. Sample KS plot of hourly and daily rainfall maxima from Kamishiiba. Results are grouped into three major categories such that Case (1) results show KS probabilities ≥ 95% for both hourly and daily sets of maxima, Case (2) results show KS probabilities < 95% for both hourly and daily sets of maxima, and Case (3) results show KS probabilities ≥ 95% for at least the hourly or the daily set of maxima.

Case (1) KS results were the ideal for this study, when both synthetic moments and maxima agree with the historical counterparts. Case (2) results were considered total failure of the NSM to model the historical maxima. Case (3) results were considered partial failures as one time scale's synthetic moments and maxima agree with its historical counterpart, while another time scale fails to model historical maxima. Both cases (2) and (3) were considered unacceptable for the purpose of this study.

While most tests yielded maxima of passing KS probabilities, the majority of cases (2) and (3) appear in the main summer months of August and September. A possible explanation for this odd result is in the mixed rainfall sources of this period. Essentially, in these months, rainfall events tend to be convective pockets of short bursts with possible medium to high intensity. It is possible within these episodes to concur with a passing of a typhoon, complicating the temporal rain cell structure of the actual cluster. It is this mixed source of rainfall that the current NSM may not be able to detect, causing it to fail in this respect. Thus, although the moments of the overall historical record is not affected (Table 3), specially in Tests I, II, and III, the varying effect of this mix of rainfall origins appeared to affect the results of the KS probabilities of these months (Table 4).

Table 4. Kolmogorov-Smirnoff Tests for synthetic maxima.

| KAMISHIIBA | JUNE | | JULY | | AUGUST | | SEPTEMBER | | OCTOBER | |
|---|---|---|---|---|---|---|---|---|---|---|
| TEST | HOURLY | DAILY | HOURLY | DAILY | HOURLY | DAILY | HOURLY | DAILY | HOURLY | DAILY |
| I | 0.998 | 0.998 | 0.995 | 0.999 | 0.941 | 0.927 | 0.953 | 0.793 | 0.997 | 0.982 |
| II | 0.994 | 0.996 | 0.999 | 0.999 | 0.874 | 0.806 | 0.894 | 0.793 | 0.884 | 0.506 |
| III | 0.884 | 0.996 | 0.999 | 0.997 | 0.830 | 0.853 | 0.972 | 0.842 | 0.995 | 0.999 |
| IV | 0.449 | 0.991 | 0.999 | 0.864 | 0.382 | 0.853 | 0.972 | 0.565 | 0.726 | 0.998 |
| V | 0.994 | 0.998 | 0.998 | 0.988 | 0.609 | 0.853 | 0.985 | 0.595 | 0.999 | 0.999 |
| VI | 0.968 | 0.999 | 0.999 | 0.994 | 0.726 | 0.853 | 0.993 | 0.842 | 0.999 | 0.948 |
| | | | | | | | | | | |
| NAHA | JUNE | | JULY | | AUGUST | | SEPTEMBER | | OCTOBER | |
| TEST | HOURLY | DAILY | HOURLY | DAILY | HOURLY | DAILY | HOURLY | DAILY | HOURLY | DAILY |
| I | 0.844 | 0.126 | 0.743 | 0.601 | 0.999 | 0.976 | 0.999 | 0.999 | 0.536 | 0.965 |
| II | 0.993 | 0.241 | 0.957 | 0.980 | 0.999 | 0.999 | 0.999 | 0.928 | 0.630 | 0.993 |
| III | 0.949 | 0.481 | 0.994 | 0.979 | 0.999 | 0.998 | 0.999 | 0.972 | 0.354 | 0.966 |
| IV | 0.982 | 0.196 | 0.957 | 0.999 | 0.991 | 0.856 | 0.996 | 0.959 | 0.920 | 0.579 |
| V | 0.936 | 0.697 | 0.992 | 0.816 | 0.999 | 0.994 | 0.999 | 0.959 | 0.927 | 0.723 |
| VI | 0.982 | 0.077 | 0.999 | 0.996 | 0.772 | 0.798 | 0.995 | 0.971 | 0.920 | 0.848 |
| | | | | | | | | | | |
| SAPPORO | JUNE | | JULY | | AUGUST | | SEPTEMBER | | OCTOBER | |
| TEST | HOURLY | DAILY | HOURLY | DAILY | HOURLY | DAILY | HOURLY | DAILY | HOURLY | DAILY |
| I | 0.997 | 0.805 | 0.995 | 0.966 | 0.998 | 0.953 | 0.993 | 0.909 | 0.603 | 1.000 |
| II | 0.891 | 0.971 | 0.998 | 0.999 | 0.998 | 0.406 | 0.964 | 0.957 | 0.999 | 0.999 |
| III | 0.947 | 0.998 | 0.973 | 0.999 | 0.999 | 0.953 | 0.999 | 0.971 | 0.999 | 0.947 |
| IV | -not applied- | | -not applied- | | -not applied- | | -not applied- | | -not applied- | |
| V | 0.999 | 0.944 | 0.999 | 0.999 | 0.999 | 0.544 | 0.999 | 0.997 | 0.999 | 0.999 |
| VI | -not applied- | | -not applied- | | -not applied- | | -not applied- | | -not applied- | |

A reformulation of the NSM, following the concept proposed by Cowpertwait (1994), may rectify the poor results for the August and September synthetic rainfall of Kamishiiba. This formulation includes "light rain cells" of long expected duration and "heavy rain cells" of short expected duration. Known to be consistent with rain field observations, the inclusion of these cells in Cowpertwait's (1994) version of NSM may be more appropriate for these mixed conditions.

The August KS probabilities within this region, the area with the highest rainfall averages among this study's study areas, are all of case (2). The KS probabilities of September on the other hand exhibit both the case (2) and case (3) results. This supports the source-mixing concept described previously, as the region experiences more rainfallin August than in September (refer to Appendix 1), although typhoons tend to be more frequent in the latter (based on the Japan Meteorological Agency). The effect of the mixed sources can thus be considered more prominent in August than in September, yielding the failing case (2) results.

In the months when the KS tests are successful, the information supplied to the parameter estimation is minimal (Test I and II), or maximal (Test V), the model generates case (1) results. This is an indication that the rainfall, when it is more consistent in origin such that the current cluster structure is valid, tends to be correlated well within a possible two-day period.

***Naha, Okinawa (1976-2002)***

The Okinawa Region is quite known to be more typhoon-frequented than the remaining parts of Japan. In the periods when the seasons change, say from Spring to Summer (May to June), as well as Summer to Autumn (October to November), it may exhibit mixes of frontal and typhoon rainfall.

Similar to the Kamishiiba results, certain months of the synthetic record generated for Naha display poor KS probabilities, spread out evenly in the case (2) and case (3) categories. However, these two months in question, July and October, are separated by three months where the model yields case (1) results quite often. Why this occurs is a crucial change from the Kamishiiba case.

Throughout the July-September period, rainfall was considered to be of a consistent source (possibly convective) and thus, the current formulation of the NSM used here was adequate for modeling maximums. The October synthetic maxima can be explained similarly to the Kamishiiba case, in which summer convective rainfall episodes are mixed with the passages of typhoons. However, as the second quarter of the year is considered the transition from the cold to the warm season of Japan (based on the Japan Meteorological Agency), the Naha Region possibly experiences a mix of the convective pockets and the passage of cold fronts in June. The effect of this passage is similar to the October results in the summer of Naha, and the August-September results of in the summer of Kamishiiba.

The mixed sources of this region are fairly less pronounced as the case (2) and (3) results are spread out evenly throughout the two months in question. However, the same reformulation proposed in the Kamishiiba case is considered appropriate here as well. With the subclassification of rain cells in the mentioned reformulation, it would be possible to accommodate the mixing effects of both June and October.

The NSM tends to be consistent in the Test II and III case for this region. This indicates that the model becomes more effective when short-term moments (6-hour and 12-hour moments) were supplied. It is possible that the general tendency of aggregated rainfall in this region, without mixed-sources of rain cells, tends to be correlated well within a possible daily period.

***Sapporo, Hokkaido (1976-2002)***

The Hokkaido Region of Japan is known to be less affected by typhoons. Indeed, only June tends to be the month when rainfall maxima origin tends to be mixed. In this region however, a case (1) type KS probability can be observed on every month, with the exception of June. In this period however, rainfall is least frequent (with an average magnitude of 2 mm per day) and was not considered a crucial month. It was thus excluded in the analysis of the results.

It was not possible for most cases to apply the NSM at the Test IV and VI condition. At this test's level of aggregation, 48 hours, the rainfall was considered practically weakly correlated. This is because the historical data in this region tends to show a change in sign of correlations (with a very small magnitude) at this aggregation level. For this reason, the Test IV and VI results were determined unreliable for the Hokkaido Region.

In this region one test condition always yielded reliable synthetic maxima, Test III in this case. It was possible to have more case (1) results in some months than others though. However, the assumption of stationarity in this region, due to less erratic typhoon episodes, is quite valid, and thus the target result in which a NSM can correctly model the maxima was obtained. For this reason, it was also considered more practical in the future to merge data in this region when rainfall occurs, rather than to model rainfall in months. In this case, it would be crucial to be aware of the precipitation conditions as snow sometimes begin in October in this region (although infrequently), for which the NSM was not conceptualized.

Upon supplying the medium-term moments of Test III (12-hour moments), the NSM becomes most effective. Although the ideal test set was similar in this region and in the previous region, these resulting tests may be considered isolated and coincidental. Further testing would prove to be appropriate at this point.

***Ideal Test Case***

In essence, the exercise showed that a range of parameters, varying by the degree of information supplied in the form of moments, should be prepared in estimating NSM parameters for the purpose at hand. It is not clear at this point that a single test set is applicable to all regions. In addition, even more sophisticated test sets can be supplied by using higher order autocorrelations or the so-called "zero-probabilities" or periods with no rainfall (see Cowpertwait, 1996).

## Concluding remarks

1. The Neyman-Scott Rectangular Pulse Rainfall Model, or NSM, was applied in this study to model rainfall maxima of certain regional rainfall in Japan: Kamishiiba (Kyushu), Naha (Okinawa), and Sapporo (Hokkaido).
2. Since the information used to estimate the parameters of NSM did not include the historical maxima, several sets of parameters, varying by the degree of information supplied (in terms of the number of historical moments) was prepared.
3. Based on these NSM parameters, random numbers different types were generated from the inverse CDF method to generate the synthetic rainfall.
4. Kyushu and Okinawa exhibited periods when rainfall is a mix of origins of frontal, convective, and typhoon type. The presence of this mix could not be accommodated in the current formulation of NSM.
5. Regardless of period or region, the moments were modeled effectively by NSM. This was the original purpose of the model conception and was treated here as a secondary objective.
6. In all cases, Kolmogorov-Smirnoff Tests were used to check whether the null hypothesis should be ignored. The KS probability of 95% was taken to disprove the null hypothesis, it proves that two samples belong to the same population.
7. The KS Test results were grouped into three cases:

(1) Both hourly and daily maxima display a KS probability $P_{KS} \geq 95\%$.
(2) Both hourly and daily maxima display a KS probability $P_{KS} < 95\%$.
(3) Only the hourly or daily maxima display a KS probability $P_{KS} \geq 95\%$.

8. As a result of the periods of "source-mixing" in 3, the KS Tests of several months appeared to have case (2) and (3) results. For Kamishiiba, these months were determined to be August to September. For Naha, these months were determined to be June and October.
9. It may be possible to capture the effect of this "mixed sources" by incorporating richer autocorrelations in the estimation procedure, or by using expressions for dry probabilities (Cowpertwait, 1996).

## References


Burlando, P. and Rosso, R., 1993. Stochastic models of temporal rainfall: Reproducibility, Estimation and Prediction of Extreme Events, Stochastic hydrology and its use in water resources systems simulation and optimization. (Marco, J. et al, ed.). NATO ASI Series. Kluwer Academic Pub.

Calenda, G. and Napolitano, F., 1999. Parameter estimation of Neyman-Scott processes for temporal point rainfall simulation. Journal of Hydrology Vol. 225 pp. 45-66.

Cowpertwait, P. 1994. A generalized point process for rainfall. Proc. R. Soc. London A 447, pp. 23-27.

Cowpertwait, P. *et al*, 1996. Neyman-Scott modeling of rainfall time series: 1. Fitting procedures for hourly and daily data. 2. Regionalization and disaggregation procedures. Journal of Hydrology Vol. 175 (1-4), pp. 17-46, 47-65.

Cox, D. and Isham, V., 1980. Point processes. Chapman and Hall, London.

Cox, D. and Isham, V., 1998. Stochastic spatial temporal models for rain. Stochastic methods in hydrology: rain, landforms and floods. (Barndorff-Nielsen, O. et al, ed). World Scientific.

Eagleson, P., 1972. Dynamics of flood frequency, Water Resources Research, vol 8, no. 4, pp. 878-898.

Favre, A. C., *et al*, 2004. Unbiased Parameter Estimation of the Neyman-Scott Model for rainfall simulation with related confidence interval. *Journal of Hydrology* Vol. 286, pp. 168-178.

Gentle, J., 1998. Random number generation and Monte Carlo methods, Springer-Verlag, New York.

Press, W. H. *et al*, 1992. Numerical Recipes in Fortran, 2nd Ed. Cambridge University Press, London.

Rodriguez-Iturbe, I. *et al*, 1987. Some models for rainfall based on stochastic point processes. Proc. R. Soc. London A 410, pp. 269-288.

**Appendix 1. Partial list of historical rainfall record moments.**

| Area/Moment | Month | | | | |
|---|---|---|---|---|---|
| **Kamishiiba** | **June** | **July** | **August** | **September** | **October** |
| **Hourly mean** | 0.7765 | 0.5826 | 0.5654 | 0.4694 | 0.1699 |
| **Hourly variance** | 2.5060 | 2.5315 | 2.5881 | 2.3487 | 1.1388 |
| **Hourly correlation, lag 1** | 0.6714 | 0.6716 | 0.6949 | 0.7207 | 0.7170 |
| **Daily mean** | 18.6354 | 13.9819 | 13.5685 | 11.2667 | 4.0786 |
| **Daily variance** | 32.9315 | 37.5997 | 37.8424 | 34.0762 | 14.2534 |
| **Daily correlation, lag 1** | 0.1516 | 0.3480 | 0.3805 | 0.3165 | 0.3324 |
| **Naha** | **June** | **July** | **August** | **September** | **October** |
| **Hourly mean** | 0.2528 | 0.2019 | 0.3326 | 0.3529 | 0.2100 |
| **Hourly variance** | 1.7434 | 1.7456 | 2.1507 | 2.4051 | 1.7380 |
| **Hourly correlation, lag 1** | 0.3812 | 0.4893 | 0.5124 | 0.5917 | 0.5706 |
| **Daily mean** | 6.0679 | 4.8459 | 7.9821 | 8.4704 | 5.0406 |
| **Daily variance** | 15.8101 | 15.9282 | 24.2758 | 29.2576 | 19.5216 |
| **Daily correlation, lag 1** | 0.1995 | 0.3765 | 0.2177 | 0.3303 | 0.2242 |
| **Sapporo** | **June** | **July** | **August** | **September** | **October** |
| **Hourly mean** | 0.0975 | 0.0975 | 0.1714 | 0.1807 | 0.1575 |
| **Hourly variance** | 0.6983 | 0.6983 | 1.0955 | 1.0267 | 0.8369 |
| **Hourly correlation, lag 1** | 0.5092 | 0.5177 | 0.6363 | 0.6844 | 0.6362 |
| **Daily mean** | 2.2593 | 2.3393 | 4.1135 | 4.3358 | 3.7802 |
| **Daily variance** | 6.9383 | 7.1159 | 14.2919 | 12.7698 | 10.0246 |
| **Daily correlation, lag 1** | 0.1010 | 0.0955 | 0.1375 | 0.1004 | 0.1206 |

# Appendix 2. Neyman-Scott Rectangular Pulse Rainfall Model Parameters.

| | Kamishiiba, Kyushu | | | | | | Naha, Okinawa | | | | | | Sapporo, Hokkaido | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **June** | **TEST I** | **TEST II** | **TEST III** | **TEST IV** | **TEST V** | **TEST VI** | **TEST I** | **TEST II** | **TEST III** | **TEST IV** | **TEST V** | **TEST VI** | **TEST I** | **TEST II** | **TEST III** | **TEST V** |
| $\lambda$ | 0.02084 | 0.02159 | 0.02200 | 0.02509 | 0.02207 | 0.02435 | 0.00977 | 0.00969 | 0.01036 | 0.00832 | 0.01032 | 0.00844 | 0.00731 | 0.00725 | 0.00733 | 0.00930 |
| $\beta$ | 0.16607 | 0.15732 | 0.16296 | 0.24876 | 0.16167 | 0.24335 | 0.08455 | 0.08683 | 0.06745 | 0.02760 | 0.08484 | 0.03254 | 0.17612 | 0.17901 | 0.18243 | 0.19137 |
| $\eta$ | 5.00000 | 2.57673 | 2.53765 | 5.00000 | 2.43229 | 5.00000 | 2.35290 | 2.37815 | 2.08139 | 1.82636 | 2.21484 | 1.82722 | 2.78975 | 2.93920 | 2.92272 | 2.16175 |
| $\mu_c$ | 49.64687 | 30.10539 | 29.43258 | 73.00657 | 28.19489 | 85.18743 | 6.20199 | 6.15552 | 4.97814 | 5.53630 | 5.32371 | 5.37889 | 15.66146 | 16.42554 | 16.17070 | 7.13131 |
| $\mu_x$ | 3.75288 | 3.07786 | 3.04320 | 2.11974 | 3.03460 | 1.87171 | 9.81589 | 10.07609 | 10.20744 | 10.02609 | 10.19558 | 10.17872 | 1.62405 | 1.64623 | 1.64390 | 3.17870 |
| **July** | **TEST I** | **TEST II** | **TEST III** | **TEST IV** | **TEST V** | **TEST VI** | **TEST I** | **TEST II** | **TEST III** | **TEST IV** | **TEST V** | **TEST VI** | **TEST I** | **TEST II** | **TEST III** | **TEST V** |
| $\lambda$ | 0.00636 | 0.00637 | 0.00633 | 0.00607 | 0.00633 | 0.00595 | 0.00185 | 0.00386 | 0.00357 | 0.00376 | 0.00386 | 0.00376 | 0.00828 | 0.00869 | 0.00883 | 0.00881 |
| $\beta$ | 0.07107 | 0.07048 | 0.07125 | 0.04955 | 0.07047 | 0.05447 | 0.01000 | 0.05950 | 0.03873 | 0.02806 | 0.05813 | 0.03687 | 0.21714 | 0.20150 | 0.20001 | 0.20088 |
| $\eta$ | 2.17691 | 2.16094 | 2.25922 | 1.17588 | 2.21982 | 1.37381 | 1.34515 | 1.70517 | 1.57853 | 1.48867 | 1.70216 | 1.59137 | 2.65969 | 2.16224 | 2.10140 | 2.10973 |
| $\mu_c$ | 44.33524 | 43.96338 | 46.10859 | 23.79551 | 45.30002 | 29.26980 | 13.93186 | 9.23424 | 8.73309 | 7.60578 | 9.23852 | 8.35625 | 10.46855 | 8.09981 | 7.77608 | 7.81367 |
| $\mu_x$ | 4.49481 | 4.49814 | 4.50912 | 4.74662 | 4.51014 | 4.59378 | 10.51604 | 9.65305 | 10.22068 | 10.51537 | 9.63448 | 10.23701 | 2.99253 | 2.99412 | 2.98387 | 2.98795 |
| **August** | **TEST I** | **TEST II** | **TEST III** | **TEST IV** | **TEST V** | **TEST VI** | **TEST I** | **TEST II** | **TEST III** | **TEST IV** | **TEST V** | **TEST VI** | **TEST I** | **TEST II** | **TEST III** | **TEST V** |
| $\lambda$ | 0.00553 | 0.00549 | 0.00549 | 0.00549 | 0.00549 | 0.00535 | 0.00658 | 0.00670 | 0.00728 | 0.00665 | 0.00718 | 0.00670 | 0.00568 | 0.00582 | 0.00569 | 0.00578 |
| $\beta$ | 0.05975 | 0.06922 | 0.06823 | 0.04840 | 0.07089 | 0.05475 | 0.10166 | 0.09704 | 0.08442 | 0.05411 | 0.09749 | 0.05041 | 0.17238 | 0.16819 | 0.17132 | 0.16794 |
| $\eta$ | 1.39585 | 1.91139 | 1.80572 | 1.06227 | 1.99791 | 1.20191 | 2.17833 | 1.95034 | 1.53775 | 1.33758 | 1.67414 | 1.05964 | 5.00000 | 4.17148 | 5.00000 | 4.43781 |
| $\mu_c$ | 30.29881 | 44.14383 | 41.61349 | 21.62012 | 46.39761 | 26.98162 | 12.69603 | 10.95158 | 8.01041 | 7.59348 | 8.87612 | 6.40984 | 42.55511 | 35.79406 | 42.13629 | 37.26940 |
| $\mu_x$ | 4.70663 | 4.45659 | 4.46598 | 5.05830 | 4.43061 | 4.70895 | 8.67195 | 8.84670 | 8.76492 | 8.80726 | 8.73239 | 8.20035 | 3.54314 | 3.43459 | 3.57516 | 3.53374 |
| **September** | **TEST I** | **TEST II** | **TEST III** | **TEST IV** | **TEST V** | **TEST VI** | **TEST I** | **TEST II** | **TEST III** | **TEST IV** | **TEST V** | **TEST VI** | **TEST I** | **TEST II** | **TEST III** | **TEST V** |
| $\lambda$ | 0.00538 | 0.00532 | 0.00545 | 0.00539 | 0.00543 | 0.00552 | 0.00401 | 0.00399 | 0.00408 | 0.00411 | 0.00412 | 0.00418 | 0.00836 | 0.00844 | 0.00826 | 0.00835 |
| $\beta$ | 0.07598 | 0.08066 | 0.07678 | 0.09199 | 0.08113 | 0.08936 | 0.05984 | 0.06836 | 0.05879 | 0.06618 | 0.06708 | 0.06426 | 0.23035 | 0.23084 | 0.22544 | 0.22452 |
| $\eta$ | 1.30860 | 1.52124 | 1.17931 | 2.18611 | 1.36007 | 1.69591 | 1.39939 | 1.52789 | 1.26258 | 1.49927 | 1.35191 | 1.34607 | 5.00000 | 5.00000 | 5.00000 | 4.48415 |
| $\mu_c$ | 26.78227 | 32.79060 | 24.29444 | 48.51414 | 28.99605 | 36.58627 | 13.83624 | 15.60732 | 12.42162 | 14.61299 | 13.54020 | 12.96417 | 49.95197 | 46.69552 | 50.34798 | 43.83992 |
| $\mu_x$ | 4.26131 | 4.09307 | 4.18351 | 3.92690 | 4.05775 | 3.94391 | 8.89091 | 8.66573 | 8.79946 | 8.80511 | 8.55639 | 8.77270 | 2.16384 | 2.29238 | 2.17229 | 2.21283 |
| **October** | **TEST I** | **TEST II** | **TEST III** | **TEST IV** | **TEST V** | **TEST VI** | **TEST I** | **TEST II** | **TEST III** | **TEST IV** | **TEST V** | **TEST VI** | **TEST I** | **TEST II** | **TEST III** | **TEST V** |
| $\lambda$ | 0.00378 | 0.00201 | 0.00372 | 0.00413 | 0.00353 | 0.00416 | 0.00409 | 0.00428 | 0.00397 | 0.00425 | 0.00412 | 0.00401 | 0.00989 | 0.01024 | 0.00976 | 0.01009 |
| $\beta$ | 0.04290 | 0.01000 | 0.03919 | 0.05503 | 0.03155 | 0.05353 | 0.09012 | 0.06984 | 0.09642 | 0.06758 | 0.07538 | 0.07824 | 0.19710 | 0.18848 | 0.19926 | 0.18948 |
| $\eta$ | 0.69869 | 0.35716 | 0.62205 | 0.84907 | 0.44722 | 0.76759 | 1.45719 | 1.14189 | 1.60154 | 1.19754 | 1.27639 | 1.36842 | 3.18336 | 2.15012 | 3.72776 | 2.30186 |
| $\mu_c$ | 8.36107 | 8.28998 | 7.63461 | 9.26514 | 6.07230 | 8.39376 | 9.24821 | 6.63743 | 10.62907 | 7.05201 | 7.58252 | 8.71386 | 26.39793 | 17.70096 | 30.00127 | 19.19743 |
| $\mu_x$ | 3.75829 | 3.65040 | 3.72499 | 3.77061 | 3.54571 | 3.73613 | 8.09587 | 8.43432 | 7.97708 | 8.39717 | 8.58266 | 8.21967 | 1.92064 | 1.86879 | 2.00476 | 1.87265 |